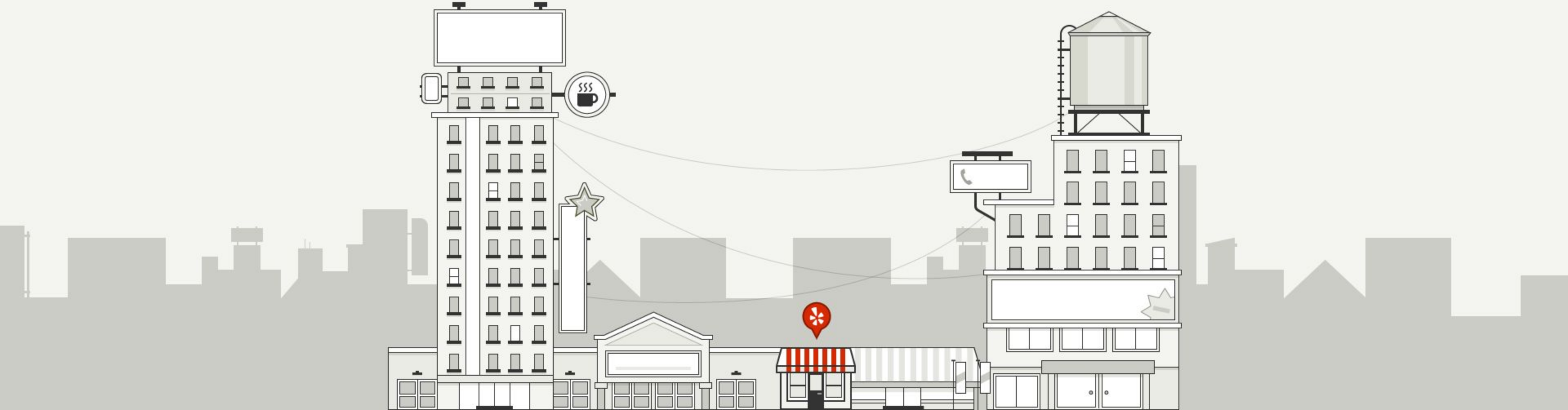
Macro vs Micro: How (and why) to write your own alerts

Megan Carney
mcarney@yelp.com/@PwnieFan



Yelp's Mission:

Connecting people with great local businesses.







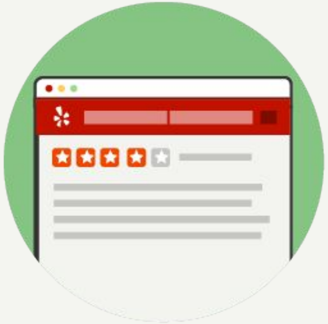


Yelp Stats

As of Q1 2016



90M



102M



70%



32

Macro Threat Intelligence

Companies that specialize in endpoint security look for patterns across their customer base, then apply those signatures or heuristics to your environment.



This is a good thing.

Micro Threat Intelligence



Analysts dedicated to your environment know what's normal (and what's not.)

This is also a good thing.

@PwnieFan

Which one do you need?

BOTH.



@PwnieFan

A Short, Simplistic History of Malware

In the beginning (1988) people wrote malware to prove it could be done.



Then they wrote malware for lulz.



Now, it’s for money or country.



Modern attackers use camouflage



- Tweak commodity grade malware to be unique enough to avoid AV
- Use your own tools against you (Powershell, SSH) to limit their footprint



@PwnieFan



@PwnieFan

Use macro and micro perspectives



“Our web proxy says that activity is suspicious, but what’s normal for our environment?”

“AV didn’t flag this executable, but it’s only installed on one of our machines. What does VirusTotal say?”

Surreal Landscape Photography by Randy Slavin

@PwnieFan

What does our DNS resolver do?

- Blocks access to ${bad} by blocking the resolution of suspect domain names
- Allows us to pull the logs about what our endpoints resolved when via the API

@PwnieFan



Spike in Blocked DNS example

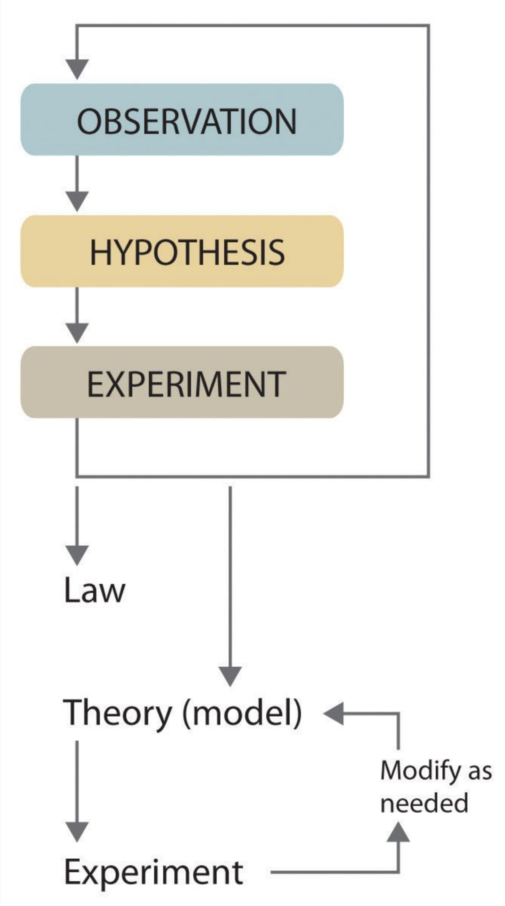


DNS resolver knows a lot about which sites are bad.



We know what our users normally resolve.

@PwnieFan



STAND BACK



I'M GOING TO TRY SCIENCE



@PwnieFan

Start by defining two groups

A group of non-infected machines (your control group)





A group of machines you know/suspect are infected.

@PwnieFan

Collect (and alert) on DNS data in ELK



```
filter:
- terms:
    opendns_data.categories:
      - "enterprise"
      - "drive-by"
      - "downloads/exploits"
      - "malware"
      - "suspicious"
      - "response"
      - "exploits"
      - "dynamic"
      - "dns"
      - "mobile"
      - "threats"
      - "botnet"
      - "phishing"
      - "high"
      - "risk"
      - "block"
```

Sample record


```
{
  "_index": "logstash-                    ",
  "_type": "opendns",
  "_id": "AVShIlVVDQqXJoZUJunf",
  "_score": null,
  "_source": {
    "@ingestionTime": "2016-05-11T18:44:19.268159Z",
    "event_time": "2016-05-11T18:29:55+00:00",
    "opendns_data": {
      "domain": "cdn.freefaits.com.",
      "external_ip": "            ,
      "response_code": "NOERROR",
      "internal_ip": "            ,
      "query_type": "1 (A)",
      "action": "Allowed",
      "identities": "                              X",
      "categories": "",
      "most_granular_identity":             "
    },
```


@PwnieFan

Researched 37 potential infections

YELP-******	not_infected	elevatedsmiles[.]c
YELP-******	not_infected	
YELP-******	not_infected	not many red flag
YELP-******	not_infected	guilt by associatic
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-		
YELP-******	not_infected	ad network crap
YELP-******	not_infected	guilt by associatic
YELP-******	not_infected	1 domain tied to a

Machine	Educated guess	Notes
YELP-******	infected	lots of illumationes[.]net domains, like *****
YELP-******	infected	http://research.zscaler.com/2015/09/compromised-wordpress-campaign-spyware.html hits three f
YELP-******	infected	lots of illumationes[.]net domains, like *****
		us
		overy[.]c
		lots of l
		n of bloc
		lso dom
YELP-******	infected	lots of illumationes[.]net domains, like *****
YELP-******	infected	lots of illumationes[.]net domains, like *****
YELP-******	infected	ycr[.]com is suspicious according to opendns and recoverlostpassword[.]com looks like bs
YELP-******	infected	t[.]f0ll0wm3[.]com
YELP-******	infected	redacted[.]redacted[.]com and h[.]redacted[.]com have redflags on VT
YELP-******	infected	www[.]colbridge[.]com is a malware site according to both VT and OpenDNS
YELP-*****	infected	looks like songwriter's website was compromised and used - VT sample 7a939e52439cb364cad0a0f

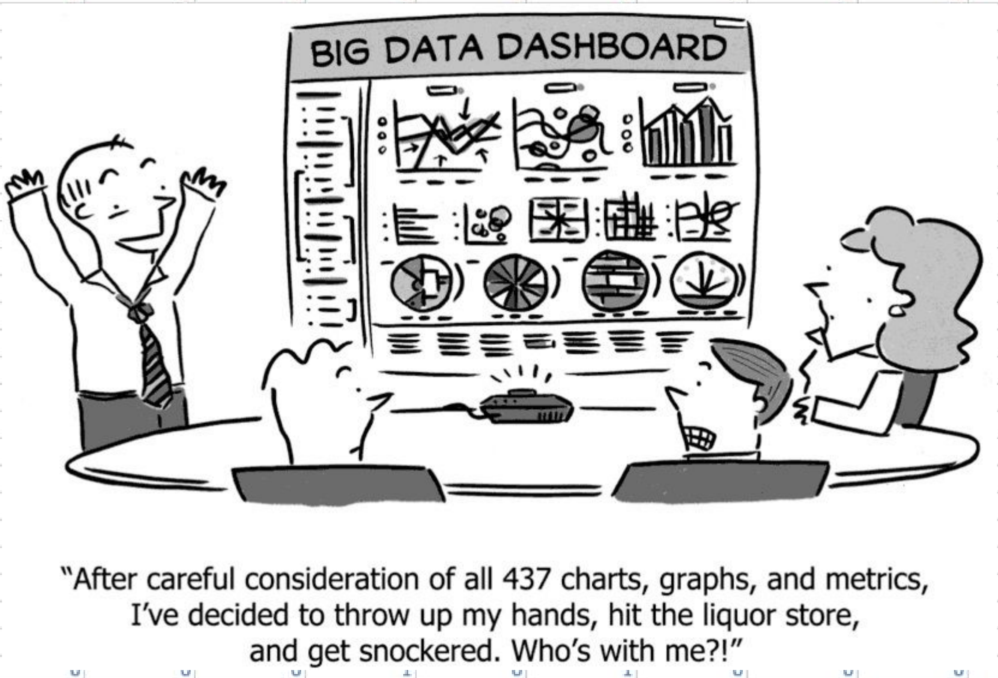
Initial research may be manual and time-consuming. It's okay to start with a small group and test more later. Just watch for over-fitting.



@PwnieFan

What differs between these groups?

Machine	Educated guess	10/7/15	10/8/15	10/9/15	10/10/15	10/11/15	10/12/15	10/13/15	10/14/15	10/15/15	10/16/15	10/17/15	10/18/15	10/19/15
YELP-******	infected	0	0	5	0	0	0	0	0	0	0	0	0	0
YELP-******	infected	5	11	0	0	0	6	5	9	0	3	0	0	0
YELP-******	infected	0	6										0	1
YELP-******	infected	0	0										0	0
YELP-******	infected	0	1										0	0
YELP-******	infected	0	0										0	0
YELP-******	infected	0	0										0	0
YELP-******	infected	0	19										0	0
YELP-******	infected	19	0										0	0
YELP-*****	infected	19	0										0	0
YELP-******	infected	19	0										0	0
YELP-******	infected	21	0										0	0
YELP-******	infected	0	29										32	41
YELP-******	infected	30	0										0	0
YELP-******	infected	30	3										0	0
YELP-******	infected	30	0										0	1
YELP-******	infected	0	0										0	0
YELP-******	infected	37	25										24	20
YELP-******	infected	0	0										0	0
YELP-*****	infected	78	0										0	0
YELP-******	not_infected	5	0										0	0
YELP-******	not_infected	5	0										0	0
YELP-******	not_infected	0	0										0	0
YELP-******	not_infected	6	2										0	2
YELP-******	not_infected	0	0										0	2
YELP-******	not_infected	0	18										0	0
YELP-******	not_infected	19	2										0	0
YELP-******	not_infected	0	0	0	26	0	0	0	0	0	0	0	0	0



@PwnieFan

How about some data visualization?

Machine	Educated guess	10/7/15	10/8/15	10/9/15	10/10/15	10/11/15	10/12/15	10/13/15	10/14/15	10/15/15	10/16/15	10/17/15	10/18/15	10/19/15
YELP-******	not_infected	5	0	0	0	0	0	0	0	0	0	0	0	0
YELP-******	not_infected	5	0	0	0	0	0	0	0	0	1	0	0	0
YELP-******	not_infected	0	0	6	0	0	0	0	0	0	0	0	0	0

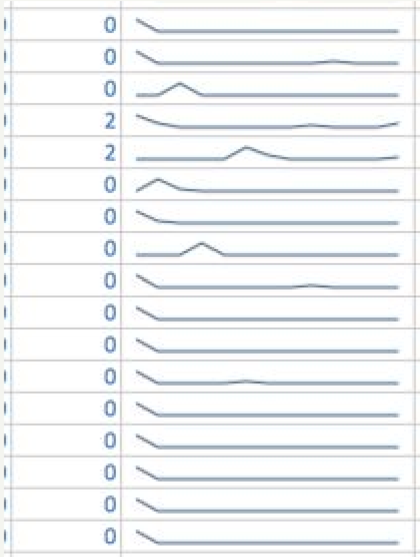
Machine	Educated guess	10/7/15	10/8/15	10/9/15	10/10/15	10/11/15	10/12/15	10/13/15	10/14/15	10/15/15	10/16/15	10/17/15	10/18/15	10/19/15
YELP-******	infected	0	0	5	0	0	0	0	0	0	0	0	0	0
YELP-******	infected	5	11	0	0	0	6	5	9	0	3	0	0	0
YELP-******	infected	0	6	1	0	0	1	0	0	0	0	0	0	1
YELP-******	infected	0	0	0	0	0	6	0	0	0	0	0	0	0
YELP-******	infected	0	1	9	0	0	4	0	0	0	6	0	0	0
YELP-******	infected	0	0	0	0	0	10	0	0	0	0	0	0	0
YELP-******	infected	0	0	18	0	0	0	10	0	0	2	0	0	0
YELP-******	infected	0	19	0	0	0	0	1	0	0	0	0	0	0
YELP-******	infected	19	0	0	0	0	0	0	1	0	0	0	0	0
YELP-*****	infected	19	0	0	0	0	0	0	0	2	0	0	0	0
YELP-******	infected	19	0	0	0	0	0	3	2	0	0	0	0	0
YELP-******	infected	21	0	2	0	0	0	0	0	0	0	0	0	0
YELP-******	infected	0	29	43	0	16	11	10	2	35	21	76	32	41
YELP-******	infected	30	0	0	0	0	0	0	0	1	2	0	0	0
YELP-******	infected	30	3	0	0	0	0	0	1	0	0	0	0	0
YELP-******	infected	30	0	0	0	0	0	4	0	0	0	0	0	1
YELP-******	infected	0	0	32	0	0	0	0	0	0	0	0	0	0
YELP-******	infected	37	25	23	26	24	24	23	24	24	23	24	24	20
YELP-******	infected	0	0	48	0	0	0	0	0	0	0	0	0	0
YELP-*****	infected	78	0	0	0	0	0	0	15	0	0	0	0	0

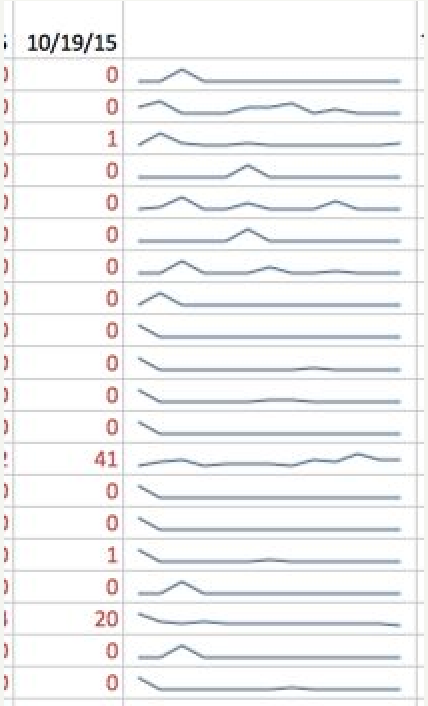


@PwnieFan

Sparklines++





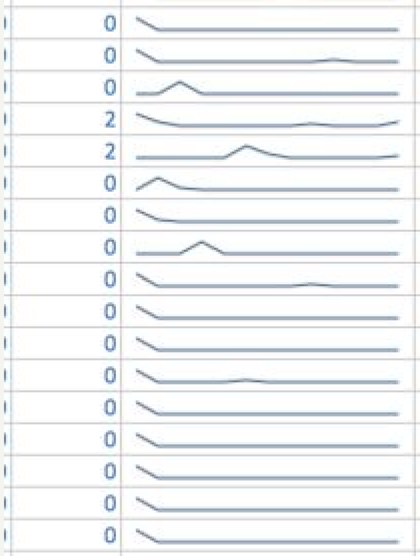




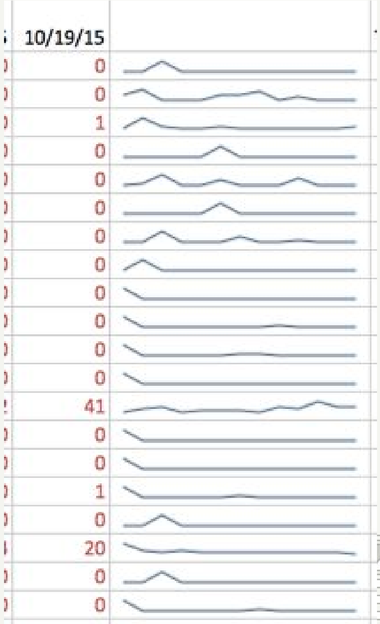
@PwnieFan

Hypothesis: A peak in blocked DNS lookups followed by low-level activity indicates infection.





0
0
0
2
2
0
0
0
0
0
0
0
0
0
0
0
0



10/19/15
0
0
1
0
0
0
0
0
0
0
0
0
41
0
0
1
0
20
0
0



@PwnieFan

Fun with Excel formulas

Machine	Educated guess	Total count	First peak >=5	Total days > 1	Total days > 2	Triggered?	Performance
YELP-******	infected	5	5	1	1	not_triggered	false_negative
YELP-******	infected	39	5	6	6	triggered	good_hit
YELP-******	infected	9	6	1	1	not_triggered	false_negative
YELP-******	infected	6	6	1	1	not_triggered	false_negative
YELP-******	infected	20	9	3	3	triggered	good_hit
YELP-******	infected	10	10	1	1	not_triggered	false_negative
YELP-******	infected	30	18	3	2	triggered	good_hit
YELP-******	infected	20	19	1	1	not_triggered	false_negative
YELP-******	infected	20	19	1	1	not_triggered	false_negative
YELP-*****	infected	21	19	2	1	triggered	good_hit
YELP-******	infected	24	19	3	2	triggered	good_hit
YELP-******	infected	23	21	2	1	triggered	good_hit
YELP-******	infected	316	29	11	10	triggered	good_hit
YELP-******	infected	33	30	2	1	triggered	good_hit
YELP-******	infected	34	30	2	2	triggered	good_hit
YELP-******	infected	35	30	2	2	triggered	good_hit
YELP-******	infected	32	32	1	1	not_triggered	false_negative
YELP-******	infected	321	37	13	13	triggered	good_hit
YELP-******	infected	48	48	1	1	not_triggered	false_negative
YELP-*****	infected	93	78	2	2	triggered	good_hit
YELP-******	not_infected	5	5	1	1	not_triggered	good_miss
YELP-******	not_infected	6	5	1	1	not_triggered	good_miss

Would my heuristic have called this machine infected?

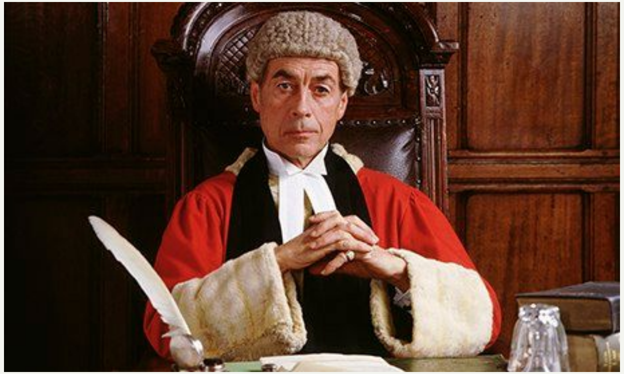
Did I call an infected machine infected? A non-infected machine infected?



@PwnieFan

How to judge an alert by the numbers

- True positive: how many infected machines did we call infected?
- True negative: how many clean machines did we call clean?



- False positive: how many clean machines did we call infected?
- False negative: how many infected machines did we call clean?

Why these numbers matter

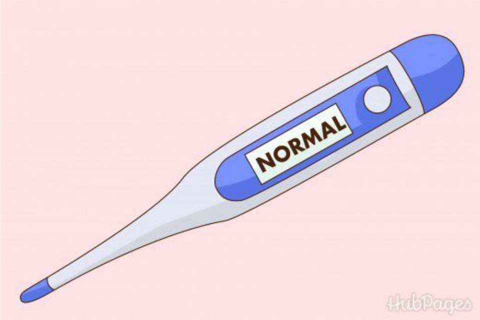
True positive





False positive

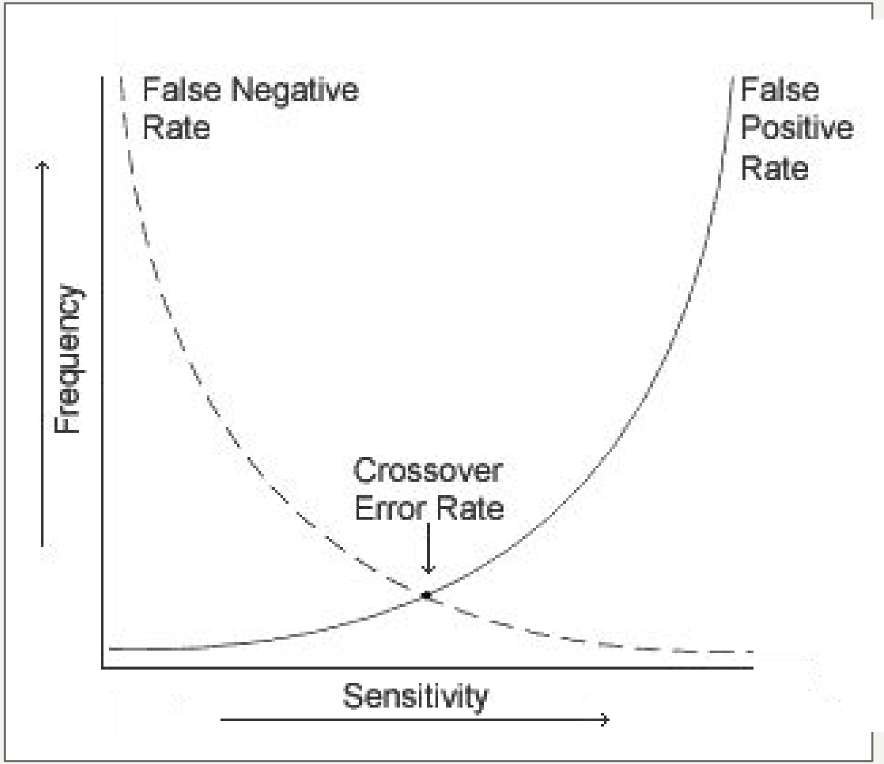
True negative





False negative

Unfortunately in general . . .



Decreasing the false negative rate will increase the false positive rate.

Finding more infections (increasing sensitivity) means triggering more false alarms.

@PwnieFan

Find your organization's balance

- How critical is the data you're protecting? Nuclear missile launch codes? Trade secrets? Inventory?
- How many people do you have for incident response? Do you automate?



@PwnieFan

After the hypothesis, experiment



What if we say an initial peak of at least 5 blocked DNS resolutions (P=5), plus at least three subsequent days with 3 or more blocked resolutions (D=3)?

True +	12	True -	10
False +	7	False -	8

Machine	Educated guess	Total count	First peak >=5	Total days > 1	Total days > 2	Triggered?	Performance					
YELP-******	infected	5	5	1	1	not_triggered	false_negative					
YELP-******	infected	39	5	6	6	triggered	good_hit		True +	12	True -	10
YELP-******	infected	9	6	1	1	not_triggered	false_negative		False +	7	False -	8
YELP-******	infected	6	6	1	1	not_triggered	false_negative					
YELP-******	infected	20	9	3	3	triggered	good_hit					

@PwnieFan

New hypothesis

Original hypothesis: P=5, D=3

True +	12	True -	10
False +	7	False -	8

New hypothesis: P=10, D=3

True +	12	True -	11
False +	6	False -	8

New new hypothesis: P=15, D=3

True +	12	True -	11
False +	6	False -	8

New new new hypothesis: P=<redacted>, D=<redacted>

True +	11	True -	13
False +	4	False -	9

How educated were those guesses?

Of the 21 machines I labeled as suspicious:

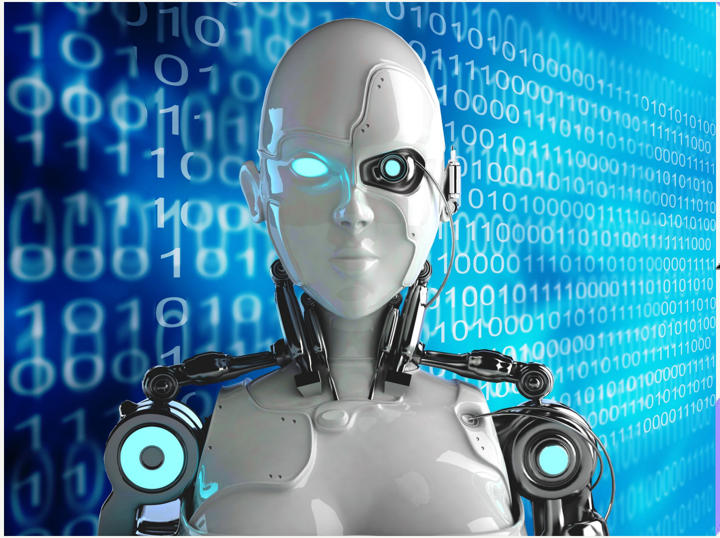
- 12 were not infected probably
- 2 were infected (confirmed via other means)
- 6 were not examined any further

If we feed those results back into our experiment . . .

True +	4	True -	18
False +	11	False -	4

@PwnieFan

False positive rate too high?



Naw, automate some analysis with osquery.

@PwnieFan

Osquery

We use osquery to periodically collect diagnostic information like:

- installed programs
- kernel extensions
- programs with listening ports
- etc

Osquery + ELK + Python

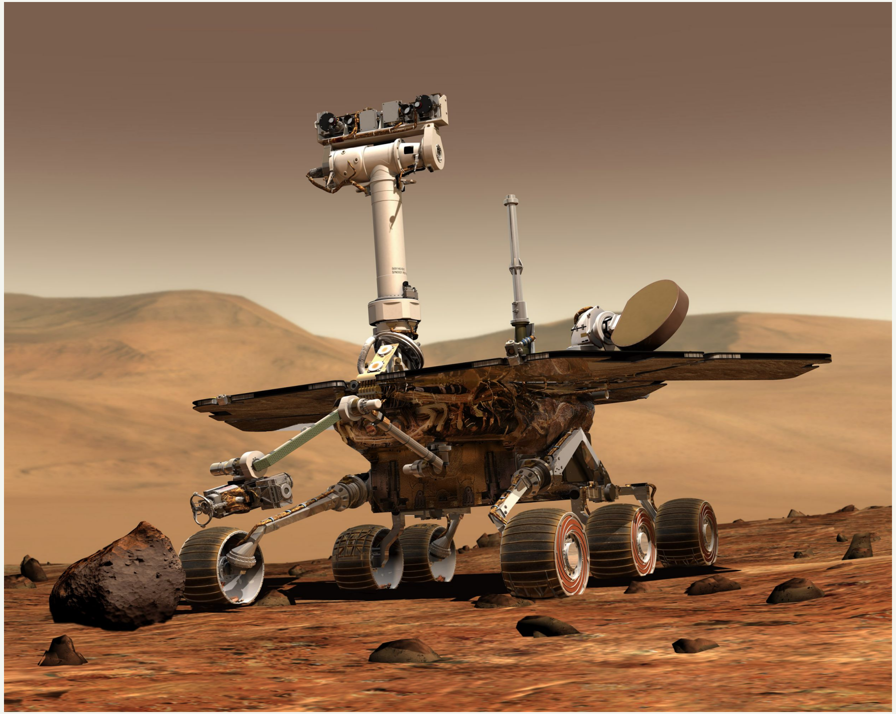


What makes this machine unique to Yelp's environment?

@PwnieFan



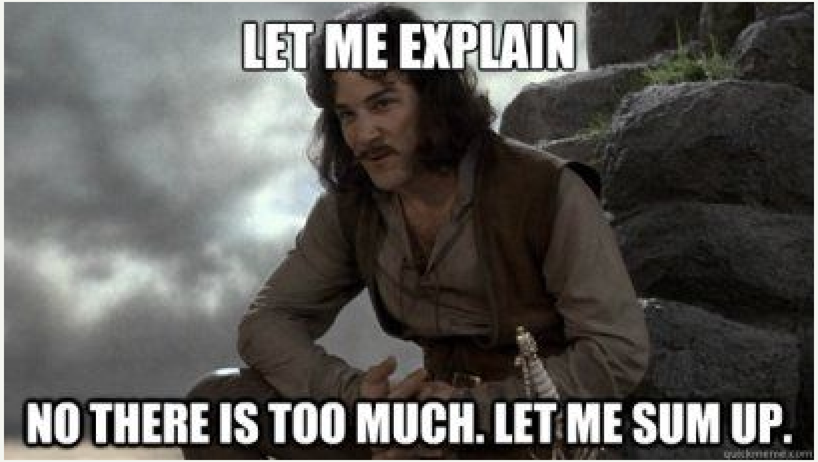
OSXcollector - now remote!



We use our package management tool to install and run OSXcollector remotely.

@PwnieFan

DNS resolver's intelligence (macro) → Local peaks in blocked resolutions (micro)



↓

SCIENCE

↓

Machines to be examined → Infections AV/IDS didn't find



@PwnieFan

Sample alert triggered

('2015-10-08', 29, {'bnk[.]annoysalbania[.]com.': 8, 'exm[.]enticingsuperpower[.]com.': 21})

('2015-10-09', 43, {'fbs[.]hearkenperturbation[.]com.': 7, 'exm[.]enticingsuperpower[.]com.': 36})

('2015-10-11', 16, {'ylu[.]lovebirdsalphabetic[.]com.': 11, 'omq[.]relievingdungeons[.]com.': 5})

('2015-10-12', 11, {'isi[.]envelopspunnet[.]com.': 11})

('2015-10-13', 10, {'isi[.]envelopspunnet[.]com.': 9, 'dlo[.]beatingsmiler[.]com.': 1})

('2015-10-14', 2, {'xwm[.]solemnpertness[.]com.': 2})

('2015-10-15', 35, {'hgh[.]filletoutdated[.]com.': 22, 'xwm[.]solemnpertness[.]com.': 13})

('2015-10-16', 21, {'hgh[.]filletoutdated[.]com.': 21})

('2015-10-17', 76, {'hgh[.]filletoutdated[.]com.': 45, 'swf[.]chequebooksbruising[.]com.': 31})

('2015-10-18', 32, {'ils[.]infernalbrazing[.]com.': 10, 'swf[.]chequebooksbruising[.]com.': 22})

('2015-10-19', 44, {'ils[.]infernalbrazing[.]com.': 41, 'elu[.]proletarianisationsinus[.]com.': 3})

('2015-10-20', 63, {'hqi[.]slugsqueals[.]com.': 51, 'elu[.]proletarianisationsinus[.]com.': 12})

Tell me more . . .

virustotal

SHA256: c15a8bad618a3e2454587c9530ce6aae95fd4c407300bb30fb76043853

Detection ratio: 12 / 56

Analysis date: 2015-10-09 19:12:06 UTC (7 months, 1 week ago)

Analysis | File detail | Additional information | Comments 0 | Vot

Antivirus	Result	
ALYac	Gen:Variant.Kazy.580380	20151009
Ad-Aware	Gen:Variant.Kazy.580380	20151009
AhnLab-V3	Trojan/Win32.Crossrider	20151009
Arcabit	Trojan.Kazy.D8DB1C	20151009
Avira (no cloud)	TR/Kazy.1513679	20151009


```
{
    "@ingestionTime": "2015-10-05T23:54:03Z",
    "columns": {
        "bundle_identifier": "com.PCvark.Advanced-Mac-Cleaner",
        "name": "Advanced Mac Cleaner.app",
        "path": "/Applications/Advanced Mac Cleaner.app"
    },
    "filter_result": "blacklisted",
    "hostIdentifier": "5",
    "kibana_link": "https:///#/dashboard/Os
}
```


@PwnieFan

Another sample alert triggered

```
('2016-01-09', 21, Counter({'standout[.]tv[.]': 21}))
('2016-01-10', 6, Counter({'ads2[.]contentabc[.]com[.]': 6}))
('2016-01-11', 5, Counter({'bttrack[.]com[.]': 2, 'cdn[.]bttrack[.]com[.]': 2,
'94982c5b634975e50103ce96082d2827[.]adsk2[.]co[.]': 1}))
('2016-01-12', 20, Counter({'ads2[.]contentabc[.]com[.]': 8, 'loadm[.]exelator[.]com
[.]': 5, 'standout[.]tv[.]': 3, 'loadus[.]exelator[.]com[.]': 2, 'secure-au[.]
imrworldwide[.]com[.]': 1, '1049theeagle[.]com[.] - Hosts malware': 1}))
('2016-01-13', 47, Counter({'ads2[.]contentabc[.]com[.]': 22, 'www[.]4chan[.]org
[.]': 14, 'sys[.]4chan[.]org[.]': 8, '4chan[.]org[.]': 2, 'cdn[.]directrev[.]com
[.]': 1}))
('2016-01-14', 2, Counter({'ads2[.]contentabc[.]com[.]': 2}))
```

@PwnieFan



Tell me more . . .

Osquery comparison script found ZipCloud.app installed:

```
00025 apps named ZipCloud.app* found
```

OSXcollector found signs of Genieo:

```
- chrome local_storage
 key: "CLOUDFLARE::http://static.genieo.com/download/download_genieo2.js"
 value: "{"url":"http://static.genieo.com/download/download_genieo2.js","contents":"var _gaq=_gaq||[];
(function()\n{var popupStyles={'default':{path:'popups/popup_mac_style.html',width:'100%',height:
'100%',margin:'0'},'with_logo':{path:'popups/popup_with_logo.html',width:'100%',height:'100%',margin:'0'}}
\n_genieo=_genieo||{};_genieo['style']=_genieo['style']||'default';var partner=_genieo['partne
r']||'genieo';var campaign=_genieo['campaign']||'';var hidden=_genieo['hidden']||false;var"
```

*ZipCloud is often bundled with undesirables

@PwnieFan

Are you a DFIR minion?



Join DFIR_MNions!

Email dfirmnions@gmail.com and tell me: who you are, why you want to join, if anyone can vouch for you.

Meetings are informal (for now)

yelp

fb.com/YelpEngineers

@YelpEngineering

engineeringblog.yelp.com

github.com/yelp